

# The impact of short term synaptic depression and stochastic vesicle dynamics on neuronal variability

Steven Reich · Robert Rosenbaum

**Abstract** Neuronal variability plays a central role in neural coding and impacts the dynamics of neuronal networks. Unreliability of synaptic transmission is a major source of neural variability: synaptic neurotransmitter vesicles are released probabilistically in response to presynaptic action potentials and are recovered stochastically in time. The dynamics of this process of vesicle release and recovery interacts with variability in the arrival times of presynaptic spikes to shape the variability of the postsynaptic response. We use continuous time Markov chain methods to analyze a model of short term synaptic depression with stochastic vesicle dynamics coupled with three different models of presynaptic spiking: one model in which the timing of presynaptic action potentials are modeled as a Poisson process, one in which action potentials occur more regularly than a Poisson process and one in which action potentials occur more irregularly. We use this analysis to investigate how variability in a presynaptic spike train is transformed by short term depression and stochastic vesicle dynamics to determine the variability of the postsynaptic response. We find that regular presynaptic spiking increases the average rate at which vesicles are released, that the number of vesicles released over a time window is more variable for smaller time windows than larger time windows and that fast presynaptic spiking gives rise to Poisson-like variability of the postsynaptic response even when presynaptic spike times are non-Poisson. Our results complement and extend previously reported theoretical results and provide possible explanations for some trends observed in recorded data.

**Keywords** Short term depression · Synaptic variability · Fano factor

## 1 Introduction

Variability of neural activity plays an important role in population coding and network dynamics [21]. Random fluctuations in the number of action potentials emitted by a population of neurons affects the firing rate of downstream cells [50, 49]. In addition, spike count variability over both short and long timescales can impact the reliability of a rate-coded signal [16]. It is therefore important to understand how this variability is shaped by synaptic and neuronal dynamics.

Several studies examine the question of how intrinsic neuronal dynamics interact with variability in presynaptic spike timing to determine the statistics of a postsynaptic neuron's spiking response, but many of these studies do not account for dynamics and variability introduced at the synaptic level by short term synaptic depression and stochastic vesicle dynamics. Synapses release neurotransmitter vesicles probabilistically in response to presynaptic spikes and recover released vesicles stochastically over a timescale of several hundred milliseconds [65, 22]. The dynamics and variability introduced by short term depression and stochastic vesicle dynamics alter the response properties of a postsynaptic neuron [60, 2, 10, 34, 26, 48, 25, 28, 17, 19, 44, 6, 43] and therefore play an important role in information

---

This work supported by National Foundation of Science grant NSF-EMSW21-RTG0739261 and National Institute of Health grant NIH-1R01NS070865-01A1.

Steven Reich  
Department of Mathematics, University of Pittsburgh, Pittsburgh PA, USA

Robert Rosenbaum (corresponding author)  
Department of Mathematics and Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh PA, USA  
Tel.: +1412-624-8375  
Fax: +1412-624-83970  
E-mail: robertr@pitt.edu

transfer [63, 18, 24, 36, 45], neural coding [57, 12, 1, 27, 20, 32, 40] and network dynamics [55, 23, 7, 62, 56, 5]. Understanding how variability in presynaptic spike times interact with short term depression and stochastic vesicle dynamics to determine the statistics of the postsynaptic response is therefore an important goal.

In this study, we use a model of short term synaptic depression with stochastic vesicle dynamics to examine how variability in a presynaptic input is transferred to variability in the synaptic response it produces. We use the theory of continuous-time Markov chains to construct exact analytical methods for calculating the the statistics of the postsynaptic response to three different presynaptic spiking models: one model with Poisson spike arrival times, one with more regular spike arrival times, and one with more irregular spike arrival times. We find that depressing synapses shape the timescale over which neuronal variability occurs: the number of neurotransmitter vesicles released over a time interval is highly variable for shorter time windows, but less variable for longer time windows when variability is quantified using Fano factors. Additionally, we find that when presynaptic inputs are highly irregular (Fano factor greater than 1), synaptic dynamics cause a reduction in Fano factor, consistent with previous studies [26, 25, 18, 19]. On the other hand, when presynaptic input is more regular (Fano factor less than 1), synaptic dynamics often cause an increase in Fano factor. This observation suggests a mechanism through which irregular and Poisson-like variability can be sustained in spontaneously spiking neuronal networks [54, 51, 8, 9, 35, 11], which complements previously proposed mechanisms [58, 61, 52, 29, 33].

## 2 Methods

We begin by introducing the synapse model used throughout this study. We then proceed by analyzing the statistics of the synaptic response to three different input models.

### 2.1 Synapse model

A widely used model of depressing synapses [57, 2, 55, 34, 48] does not capture stochasticity in vesicle recovery and release. As a result, this model underestimates the variability of the synaptic response [19, 43]. For this reason, we use a more detailed synapse model that takes stochastic recovery times and probabilistic release into account [60, 62, 22, 43].

We consider a presynaptic neuron with spike train  $I(t) = \sum_j \delta(t - t_j)$  that makes  $M$  functional contacts

onto a postsynaptic cell. Here,  $t_j$  is the time of the  $j$ th presynaptic action potential. Define  $m(t)$  to be the number of contacts with a readily releasable neurotransmitter vesicle at time  $t$  (so that  $0 \leq m(t) \leq M$ ). For simplicity, we assume that each contact can release at most one neurotransmitter vesicle in response to a presynaptic spike. When a presynaptic spike arrives, each contact with a releasable vesicle releases its vesicle independently with probability  $p_r$ . After releasing a vesicle, a synaptic contact enters a refractory period during which it is unavailable to release a vesicle again until it recovers by replacing the released vesicle. The recovery time at a single contact is modeled as a Poisson process with rate  $1/\tau_u$ . Equivalently, the duration of the refractory period is exponentially distributed with mean  $\tau_u$ .

Define  $w_j$  to be the number of contacts that release a vesicle in response to the presynaptic spike at time  $t_j$  (so that  $0 \leq w_j \leq m(t_j^-) \leq M$  where  $m(t_j^-) = \lim_{t \rightarrow t_j^-} m(t)$ ). The synaptic response is quantified by the marked point process

$$x(t) = \sum_j w_j \delta(t - t_j).$$

Since the signal observed by the postsynaptic cell is determined by  $x(t)$ , we quantify synaptic response statistics in terms of the statistics of  $x(t)$  in our analysis. The process  $x(t)$  can be convolved with a post-synaptic response kernel to obtain the conductance induced on the postsynaptic cell [43]. The effects of this convolution on response statistics is well understood [53], so we do not consider it here.

This model can be described more precisely using the equation [43]

$$dm(t) = -dN_x(t) + dN_u(t)$$

where  $dN_u(t) = u(t)dt$  is the increment of an inhomogeneous Poisson process with instantaneous rate that depends on  $m(t)$  through  $\langle dN_u(t) | m(t) \rangle = dt(M - m(t))/\tau_u$  (here,  $\langle \cdot | \cdot \rangle$  denotes conditional expectation),  $N_x(t) = \int_0^t x(s)ds$  is the number of vesicles released up to time  $t$ , and each  $w_j$  is a binomial random variable with mean  $p_r m(t_j)$  and variance  $m(t_j)p_r(1 - p_r)$ .

### 2.2 Statistical measures of the presynaptic spike train and the synaptic response

We focus on steady state statistics in this article, and therefore assume that the presynaptic spike trains are stationary and that the synapses have reached statistical equilibrium. The intensity of a presynaptic spike train is quantified by the mean presynaptic firing rate,

$$r_{in} = \langle I(t) \rangle = \langle N_I(T) / T \rangle$$

where  $\langle \cdot \rangle$  denotes the expected value and

$$N_I(T) = \int_0^T I(s)ds$$

represents the number of spikes in the time interval  $[0, T]$ . Temporal correlations in the presynaptic spike times are quantified by the auto-covariance,

$$R_{\text{in}}(\tau) = \text{cov}(I(t), I(t + \tau)),$$

and the variability in the presynaptic spike train is quantified by its Fano factor,

$$F_{\text{in}}(T) = \frac{\text{var}(N_I(T))}{Tr_{\text{in}}}.$$

For much of this work, we will focus on Fano factors over large time windows which, through a slight abuse of notation, we denote by  $F_{\text{in}} = \lim_{T \rightarrow \infty} F_{\text{in}}(T)$ . To compute Fano factors, we will often exploit their relationship to auto-covariance functions,

$$F_{\text{in}}(T) = \frac{1}{r_{\text{in}}} \int_{-T}^T R_{\text{in}}(\tau)(1 - |\tau|/T)d\tau \quad (1)$$

and

$$F_{\text{in}} = \frac{1}{r_{\text{in}}} \int_{-\infty}^{\infty} R_{\text{in}}(\tau)d\tau. \quad (2)$$

The statistics of the synaptic response,  $x(t)$ , are defined analogously to the statistics of  $I(t)$ . The steady state rate of vesicle release is defined as

$$r_x = \langle x(t) \rangle = \langle N_x(T)/T \rangle$$

where  $N_x(T) = \int_0^T x(s)ds$  represents the number of vesicles released in the time interval  $[0, T]$ . Temporal correlations in the synaptic response are quantified by the auto-covariance,

$$R_x(\tau) = \text{cov}(x(t), x(t + \tau))$$

and response variability is quantified by the Fano factor of the number of vesicles released,

$$F_x(T) = \frac{\text{var}(N_x(T))}{Tr_x}.$$

As above, we define  $F_x = \lim_{T \rightarrow \infty} F_x(T)$  and note that

$$F_x(T) = \frac{1}{r_{\text{in}}} \int_{-T}^T R_x(\tau)(1)d\tau \quad \text{and} \quad F_x = \int_{-\infty}^{\infty} R_x(\tau)d\tau. \quad (3)$$

### 2.3 Model analysis with Poisson presynaptic inputs

We first consider a homogeneous Poisson input,  $I(t)$ , with rate  $r_{\text{in}}$ . The input auto-covariance for this model is given by  $R_{\text{in}}(\tau) = r_{\text{in}}\delta(\tau)$  and the Fano factor is given by  $F_{\text{in}}(T) = 1$  for any  $T > 0$ . The mean rate of vesicle release for this model is given by

$$r_x = \frac{Mp_r r_{\text{in}}}{p_r r_{\text{in}} \tau_u + 1}$$

which saturates to  $M/\tau_u$  for large presynaptic rates,  $r_{\text{in}}$ . A closed form approximations to the auto-covariance function of the response for this Poisson input model are derived in [43,36] (see also [17]) and consist of a sum of a delta function and an exponential,

$$R_x(\tau) = Dr_x\delta(\tau) - Er_x e^{-|\tau|/\tau_0}, \quad (4)$$

where the mass of the delta function is given by

$$D = \frac{2p_r(r_{\text{in}}\tau_u + M - 1) + 2 - p_r^2 r_{\text{in}}\tau_u}{(2 - p_r)p_r r_{\text{in}}\tau_u + 2} > 0, \quad (5)$$

the timescale of the exponential decay is given by

$$\tau_0 = \frac{\tau_u}{1 + p_r r_{\text{in}}\tau_u},$$

and the peak of the exponential is given by

$$E = \frac{p_r r_{\text{in}}((M - 2)p_r + 2)\tau_u + 2(M - 1)p_r + 2}{M(2 - p_r)p_r r_{\text{in}}\tau_u + 2M} r_x. \quad (6)$$

It can easily be checked that  $E > 0$  whenever  $M \geq 1$ ,  $0 \leq p_r \leq 1$ ,  $r_{\text{in}} > 0$ , and  $\tau_u > 0$  so that the peak of the exponential in (4) is negative. For finite  $T$ , the Fano factor,  $F_x(T)$ , is given by

$$F_x(T) = D - 2E\tau_0 - \left(e^{-\frac{T}{\tau_0}} - 1\right) \frac{2E\tau_0^2}{T} \quad (7)$$

and, in the limit of large  $T$ ,

$$F_x = D - 2E\tau_0. \quad (8)$$

To test the accuracy of these approximations, exact solutions can be found numerically using standard methods for the analysis of continuous-time Markov chains, as described for alternate input models below. This analysis is a special case of the analysis for the “regular” input model described below that is achieved by taking  $\theta = 1$ . Alternatively, exact numerical results can be achieved by taking  $r_s = r_f$  for the “irregular” input model. In figures showing results for the Poisson input model, we plot the closed form approximations described above along with exact numerical results obtained using the regular input model with  $\theta = 1$ .

## 2.4 Model analysis with irregular presynaptic inputs

Spike trains measured *in vivo* often exhibit irregular spiking statistics indicated by Fano factors larger than 1 [4, 15, 3, 11]. To describe the synaptic response to irregular inputs, we use a model of presynaptic spiking in which the instantaneous rate of the presynaptic spike train,  $I(t)$ , randomly switches between two values,  $r_s$  and  $r_f > r_s$ , representing a slow spiking state and a fast spiking state. The time spent in the slow state before transitioning to the fast state is exponentially distributed with mean  $\tau_s$ . Likewise, the amount of time spent in the fast state before switching to the slow state is exponentially distributed with mean  $\tau_f$ . Transition times are independent from one another and from the spiking activity. Between transitions, spikes occur as a Poisson process.

To find  $r_{\text{in}}$ ,  $R_{\text{in}}(\tau)$ , and  $F_{\text{in}}$ , we represent this model as a doubly stochastic Poisson process. Define  $r(t) \in \{r_s, r_f\}$  to be the instantaneous firing rate at time  $t$ . Then  $r(t)$  is a continuous time Markov chain [31] on the state space  $\Gamma = (r_s, r_f)$  with infinitesimal generator matrix

$$A = \begin{bmatrix} -1/\tau_s & 1/\tau_s \\ 1/\tau_f & -1/\tau_f \end{bmatrix}.$$

Clearly,  $r(t)$  spends a proportion  $\tau_s/(\tau_s + \tau_f)$  of its time in the slow state (defined by  $r(t) = r_s$ ) and a proportion  $\tau_f/(\tau_s + \tau_f)$  of its time in the fast state (defined by  $r(t) = r_f$ ). This gives a steady-state mean firing rate of

$$r_{\text{in}} = \frac{r_s \tau_s + r_f \tau_f}{\tau_s + \tau_f}.$$

At non-zero lags ( $\tau \neq 0$ ), the auto-covariance of a doubly stochastic Poisson process is the same as the auto-covariance of  $r(t)$  [43], which we can compute using techniques for analyzing continuous time Markov chains. For  $\tau > 0$ , we have

$$\begin{aligned} \langle r(t)r(t+\tau) \rangle &= r_s \Pr(r(t) = r_s) \langle r(t+\tau) | r(t) = r_s \rangle \\ &\quad + r_f \Pr(r(t) = r_f) \langle r(t+\tau) | r(t) = r_f \rangle \\ &= \frac{r_s \tau_s}{\tau_s + \tau_f} \langle r(t+\tau) | r(t) = r_s \rangle \quad (9) \\ &\quad + \frac{r_f \tau_f}{\tau_s + \tau_f} \langle r(t+\tau) | r(t) = r_f \rangle \end{aligned}$$

where  $\langle \cdot | \cdot \rangle$  denotes conditional expectation and

$$\begin{aligned} \langle r(t+\tau) | r(t) = r_s \rangle &= r_s \Pr(r(t+\tau) = r_s | r(t) = r_s) \\ &\quad + r_f (1 - \Pr(r(t+\tau) = r_s | r(t) = r_s)). \end{aligned}$$

The probability in this expression can be written in terms of an exponential of the generator matrix,  $A$ , and

then calculated explicitly to obtain

$$\begin{aligned} \Pr(r(t+\tau) = r_s | r(t) = r_s) &= \left[ e^{A^T \tau} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right]_1 \\ &= \frac{\tau_f e^{-\frac{\tau}{\tau_f} - \frac{\tau}{\tau_s}} + \tau_s}{\tau_f + \tau_s} \end{aligned}$$

where  $[v]_k$  denotes the  $k$ th component of a vector,  $v$ . An identical calculation can be performed to obtain an analogous expression for  $\langle r(t+\tau) | r(t) = r_f \rangle$ . Combining these with Eq. (9) gives

$$\langle r(t)r(t+\tau) \rangle = \frac{\tau_f \tau_s (r_f - r_s)^2}{(\tau_f + \tau_s)^2} e^{-\frac{\tau}{\tau_s} - \frac{\tau}{\tau_f}} + r_{\text{in}}^2.$$

For positive  $\tau$ , we have  $R_{\text{in}}(\tau) = \langle r(t)r(t+\tau) \rangle - r_{\text{in}}^2$ . As with all stationary point processes  $R_{\text{in}}(\tau) = R_{\text{in}}(-\tau)$  and  $R_{\text{in}}(\tau)$  has a Dirac delta function with mass  $r_{\text{in}}$  at the origin [14]. Thus, the auto-covariance of  $I(t)$  is given by

$$R_{\text{in}}(\tau) = r_{\text{in}} \delta(\tau) + \frac{\tau_f \tau_s (r_f - r_s)^2}{(\tau_f + \tau_s)^2} e^{-\frac{|\tau|}{\tau_s} - \frac{|\tau|}{\tau_f}}. \quad (10)$$

For finite  $T$ , the Fano factor,  $F_{\text{in}}(T)$ , can be computed using Eqs. (1) and (10). In the limit of large  $T$ , we can use Eqs. (2) and (10) to obtain a closed form expression,

$$F_{\text{in}} = 1 + \frac{2\tau_f^2 \tau_s^2 (r_f - r_s)^2}{(\tau_f + \tau_s)^2 (r_f \tau_f + r_s \tau_s)}. \quad (11)$$

Poisson spiking is recovered by setting  $r_f = r_s$ ,  $\tau_f = 0$ , or  $\tau_s = 0$ . For any other parameter values (i.e., when  $r_f \neq r_s$  and  $\tau_f, \tau_s > 0$ ), it follows from Eq. (11) that  $F_{\text{in}}(T) > 1$  for any  $T$ . Therefore this input model, hereafter referred to as the “irregular spiking” model, represents spiking that is more irregular than a Poisson process.

The analysis in [43] used to derive closed form expressions for the response statistics with Poisson inputs cannot easily be generalized to derive expressions with non-Poisson inputs like those considered here. Instead, we analyze the synaptic response for the irregular input model using techniques for analyzing continuous time Markov chains. First note that the process  $b(t) = (m(t), r(t))$  is a continuous-time Markov chain on the discrete state space  $\{0, 1, \dots, M\} \times \{r_s, r_f\}$ . Here,  $m(t)$  denotes the size of readily releasable pool and  $r(t)$  represents the instantaneous presynaptic rate (which switches between  $r_s$  and  $r_f$ ). We enumerate all  $2(M+1)$  elements of this state space and denote the  $j$ th element of this enumeration as  $\Gamma_j = (m_j, r_j)$  for  $j = 1, \dots, 2(M+1)$ .

The infinitesimal generator,  $B$ , of  $b(t)$  is a  $2(M+1) \times 2(M+1)$  matrix with off-diagonal terms defined by the instantaneous transition rates,

$$B_{j,k} = \lim_{h \rightarrow 0} \frac{1}{h} \Pr(b(t+h) = \Gamma_k | b(t) = \Gamma_j), \quad j \neq k \quad (12)$$

and with diagonal terms chosen so that the rows sum to zero:  $B_{j,j} = -\sum_{k \neq j} B_{j,k}$  [31].

To fill the matrix  $B$ , we consider each type of transition that the process  $b(t)$  undergoes. Vesicle recovery events occur at the instantaneous rate  $(M - m(t))/\tau_u$  and increment the value of  $m(t)$  by one vesicle. Therefore

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(b(t+h) = (m+1, r) | b(t) = (m, r)) = \frac{M-m}{\tau_u}$$

for  $m \in \{0, \dots, M-1\}$  and  $r \in \{r_s, r_f\}$ . Vesicle release events occur at the instantaneous rate  $r(t)$  and decrement the value of  $m(t)$  by a random amount  $k$  with a binomial distribution so that

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(b(t+h) = (m-k, r) | b(t) = (m, r)) = r \frac{m!}{(m-k)!} p_r^k (1-p_r)^{m-k}$$

for  $m \in \{1, \dots, M\}$ ,  $k \in \{0, \dots, m\}$ , and  $r \in \{r_s, r_f\}$ . The value of  $r(t)$  switches from  $r_s$  to  $r_f$  with instantaneous rate  $1/\tau_s$  so that

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(b(t+h) = (m, r_f) | b(t) = (m, r_s)) = \frac{1}{\tau_s}$$

and, similarly,

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(b(t+h) = (m, r_s) | b(t) = (m, r_f)) = \frac{1}{\tau_f}.$$

These four transition types account for all of the transitions that  $b(t)$  undergoes. They can be used to fill the off-diagonal terms of the matrix  $B$ . The diagonal terms are then filled to make the rows sum to zero, as discussed above.

Once a the infinitesimal generator matrix,  $B$ , is obtained, the probability distribution of  $b(t)$  given an initial distribution  $p(0)$  is given by

$$p(t) = e^{tB^T} p(0).$$

The stationary distribution,  $p_0$ , of  $b(t)$  is given by the vector in the one-dimensional null space of  $B$  with elements that sum to one [31].

The instantaneous rate of vesicle release, conditioned on the current state of  $r(t)$  and  $m(t)$ , is given by

$$\langle x(t) | r(t) = r, m(t) = m \rangle = rp_r m.$$

Averaging over  $r$  and  $m$  in the steady state gives

$$r_x = \sum_{j=1}^{2(M+1)} [p_0]_j r_j p_r m_j$$

where  $[\cdot]_j$  denotes the  $j$ th element. The auto-covariance,  $R_x(\tau)$ , has a Dirac delta function at  $\tau = 0$ . We separate this delta function from the continuous part by writing  $R_x(\tau) = A_x \delta(\tau) + R_x^+(\tau)$  where  $R_x^+(\tau)$  is a continuous function. The area of the delta function can be found by conditioning on the current state of  $r(t)$  in the steady state to get

$$\begin{aligned} A_x &= \lim_{t \rightarrow \infty} \langle x(t)^2 dt \rangle \\ &= \lim_{t \rightarrow \infty} \langle dN_x^2(t)/dt \rangle \\ &= \sum_{j=1}^{2(M+1)} r_j [p_0]_j \lim_{k \rightarrow \infty} \langle w_k^2 | m(t_k^-) = m_j \rangle \end{aligned} \quad (13)$$

where  $w_k$  is the number of vesicles released by the  $k$ th presynaptic spike. Conditioned on the size,  $m(t_k^-)$ , of the readily releasable pool immediately before the presynaptic spike arrives,  $w_k$  has a binomial distribution with second moment,

$$\lim_{k \rightarrow \infty} \langle w_k^2 | m(t_k^-) = m_j \rangle = m_j p_r (1-p_r) + m_j^2 p_r^2$$

which can be substituted into Eq. (13) to calculate  $A_x$ .

All that remains is to calculate the continuous part,  $R_x^+(\tau)$ , of  $R_x(\tau)$ . First note that, for  $\tau > 0$ ,

$$\begin{aligned} &\lim_{t \rightarrow \infty} \langle x(t)x(t+\tau) \rangle \\ &= \sum_{i,j=1}^{2(M+1)} [p_0]_i \Pr(b(t+\tau) = \Gamma_j | b(t) = \Gamma_i) \\ &\times \langle dN_x(t)dN_x(t+\tau) | b(t) = \Gamma_i, b(t+\tau) = \Gamma_j \rangle / dt^2. \end{aligned} \quad (14)$$

The second term in Eq. (14) can be computed as

$$\Pr(b(t+\tau) = \Gamma_j | b(t) = \Gamma_i) = [e^{\tau B^T} \mathbf{e}_i]_j = [e^{\tau B^T}]_{j,i}$$

where  $\mathbf{e}_i$  is the  $2(M+1) \times 1$  vector whose  $i$ th element is 1 and all other elements are zero, which represents an initial distribution concentrated at  $\Gamma_i$ . The last term in Eq. (14) is given by

$$\langle dN_x(t)dN_x(t+\tau) | b(t) = \Gamma_i, b(t+\tau) = \Gamma_j \rangle / dt^2 = r_i r_j p_r^2 m_i m_j.$$

Finally,  $R_x^+(\tau) = \lim_{t \rightarrow \infty} \langle x(t)x(t+|\tau|) \rangle - r_x^2$  for  $\tau \neq 0$  so that

$$R_x(\tau) =$$

$$A_x \delta(\tau) - r_x^2 + \sum_{i,j=1}^{2(M+1)} [p_0]_i \left[ e^{|\tau| B^T} \right]_{j,i} r_i r_j p_r^2 m_i m_j$$

which can be computed efficiently using matrix multiplication. The response Fano factor,  $F_x$ , can then be found by integrating  $R_x(\tau)$  according to Eqs. (1) and (2).

## 2.5 Model analysis with regular presynaptic inputs

We now consider a spiking model that gives Fano factors smaller than 1 and therefore spike trains that are more regular than Poisson processes. We achieve this by defining a renewal process with gamma-distributed interspike intervals (ISIs). Such a process can be obtained by first generating a Poisson process,  $\sum_k \delta(t - s_k)$  with rate  $r = \theta r_{\text{in}}$  for some positive integer  $\theta$ , then keeping only every  $\theta$ th spike to build the spike train  $I(t)$ . More precisely, the first spike of the gamma process is obtained by choosing an integer,  $k$ , uniformly from the set  $\{1, \dots, \theta\}$  and defining  $t_1 = s_k$ . The remaining spikes are defined by  $t_{j+1} = s_{j\theta+k}$  to obtain the stationary renewal process,  $I(t) = \sum_j \delta(t - t_j)$  [13].

Clearly, this process has rate  $r_{\text{in}}$  since the original Poisson process has rate  $\theta r_{\text{in}}$  and a proportion  $1/\theta$  of these spikes appear in  $I(t)$ . The auto-covariance is given by [42]

$$R_{\text{in}}(\tau) = r_{\text{in}}\delta(\tau) + r_{\text{in}} \left( \sum_{k=1}^{\infty} f_k(\tau) - r_{\text{in}} \right). \quad (15)$$

where

$$f_k(t) = \frac{t^{k\theta-1} (\theta r_{\text{in}})^{k\theta} e^{-\theta r_{\text{in}} t}}{(k\theta - 1)!}$$

is the density of the waiting time between the first spike and the  $(k+1)$ st spike (i.e., the duration of  $k$  consecutive ISIs).

For finite  $T$ , the Fano factor,  $F_{\text{in}}(T)$ , can be computed using Eqs. (1) and (15). In the limit of large  $T$ , we can use Eq. (2) or use the fact that, for renewal processes,  $F_{\text{in}} = \text{var}(\text{ISI})/\langle \text{ISI} \rangle^2$  where  $\text{var}(\text{ISI}) = 1/(r_{\text{in}}^2 \theta)$  is the variance and  $\langle \text{ISI} \rangle = 1/r_{\text{in}}$  is the mean of the gamma distributed ISIs [13]. This gives

$$F_{\text{in}} = \frac{1}{\theta}.$$

Poisson spiking is recovered by setting  $\theta = 1$ . When  $\theta > 1$ , we have that  $F_{\text{in}}(T) < 1$  for any  $T$ . Therefore this model, hereafter referred to as the “regular” input model, represents spiking that is more regular than a Poisson process.

The synaptic response with the regular input model can be analyzed using methods similar to those used for the irregular model. We introduce an auxiliary process,  $q(t)$ , that transitions sequentially through the state space  $\{1, \dots, \theta\}$ . Once reaching  $\theta$ ,  $q(t)$  transitions back to

state 1. Transitions occur as a Poisson process with rate  $\theta r_{\text{in}}$ . The waiting times between transitions from  $q = \theta$  to  $q = 1$  are gamma distributed. Thus, to recover the regular input model, we specify that each transition from  $1 = \theta$  to  $q = 1$  represents a single presynaptic spike. The process  $g(t) = (m(t), q(t))$  is then a continuous time Markov chain on the discrete state space  $\{1, \dots, \theta\} \times \{0, \dots, M\}$ . We enumerate all  $\theta(M+1)$  elements of this space and denote the  $j$ th element as  $\Gamma_j = (m_j, q_j)$  for  $j = 1, \dots, \theta(M+1)$ .

The infinitesimal generator,  $G$ , which is a  $\theta(M+1) \times \theta(M+1)$  matrix is defined analogously to the matrix  $B$  in Eq. (12) above. The elements of  $G$  can be filled using the following transition probabilities. As for the irregular input model, vesicle recovery occurs as a Poisson process with rate  $(M - m(t))/\tau_u$  so that

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(g(t+h) = (m+1, q) | g(t) = (m, q)) = \frac{M - m}{\tau_u}$$

for  $m = 0, \dots, M$  and  $q = 1, \dots, \theta$ . Transitions that increment  $q(t)$  occur with instantaneous rate,  $\theta r_{\text{in}}$  so that

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(g(t+h) = (m, q+1) | g(t) = (m, q)) = \theta r_{\text{in}}$$

for  $q = 1, \dots, \theta-1$  and  $m = 0, \dots, M$ . The only other transitions are those from  $q(t) = \theta$  to  $q(t) = 1$ , which represent a presynaptic spike and are therefore accompanied by a release of vesicles. The transitions contribute the following,

$$\lim_{h \rightarrow 0} \frac{1}{h} \Pr(g(t+h) = (1, m-k) | g(t) = (\theta, m)) =$$

$$\theta r_{\text{in}} \frac{m!}{(m-k)!} p_r^k (1-p_r)^{m-k}$$

for  $m \in \{0, 1, \dots, M\}$  and  $k \in \{0, \dots, m\}$ . These transition rates can be used to fill the off-diagonal terms of the matrix  $G$ . The diagonal terms are then filled so that the rows sum to zero. The stationary distribution,  $p_0$ , of  $g(t) = (m(t), q(t))$  is given by the vector in the one-dimensional null space of  $G$  with elements that sum to one.

A proportion  $[p_0]_{\gamma(k)}$  of time is spent in state  $(m(t), q(t)) = (k, \theta)$  where  $\gamma(k)$  represents the index of the element  $(k, \theta)$  in the enumeration chosen for  $\Gamma$  (i.e., the index,  $j$ , at which  $\Gamma_j = (k, \theta)$ ). In that state, the transition to  $q(t) = 1$  occurs with instantaneous rate  $\theta r_{\text{in}}$  and releases average of  $p_r m(t)$  vesicles. Thus, the mean rate of vesicle release is given by

$$r_x = \sum_{k=1}^M \theta r_{\text{in}} p_r k [p_0]_{\gamma(k)}.$$

As above, we separate the auto-covariance into a delta function and a continuous part by writing  $R_x(\tau) =$

$A_x \delta(\tau) + R_x^+(\tau)$  where  $R_x^+(\tau)$  is a continuous function. The area of the delta function at the origin is given by

$$A_x = \sum_{k=1}^M \theta r_{\text{in}}[p_0]_{\gamma(k)} (kp_r(1 - p_r) + k^2 p_r^2)$$

by an argument identical to that used for the irregular input model above. Also by a similar argument used for the irregular input model, we have that

$$R_x(\tau) = A_x \delta(\tau) - r_x^2 + \theta^2 r_{\text{in}}^2 \sum_{k,l=1}^M [p_0]_{\gamma(k)} \left[ e^{|\tau|G^T} \right]_{\gamma(l), \gamma(k)} p_r^2 k l.$$

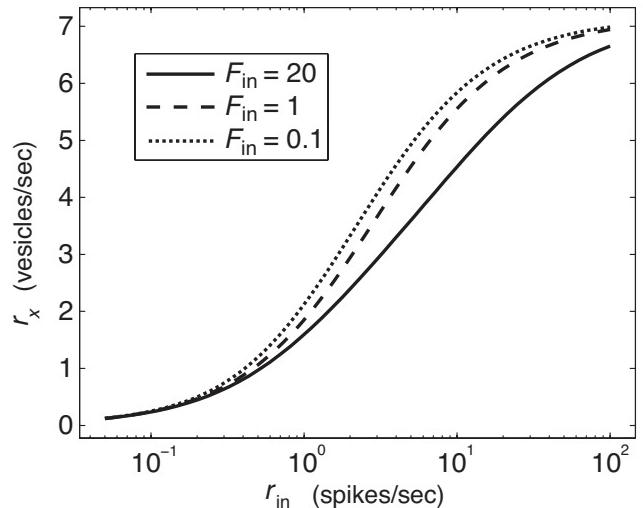
## 2.6 Parameters used in figures

Theoretical results are obtained for arbitrary parameter values, but for all figures we use a set of parameter values that are consistent with experimental studies. For synaptic parameters, we use  $\tau_u = 700$  ms and  $p_r = 0.5$  consistent with measurements of short term depression in pyramidal-to-pyramidal synapses in the rat neocortex [57,22]. We also choose  $M = 5$  which is within the range observed in several cortical areas [6].

The Poisson presynaptic input model is determined completely by its firing rate and the regular input model is determined completely by its firing rate and Fano factor. Presynaptic firing rates and Fano factors are reported on the axes or captions of each figure. The irregular input model has four parameters that determine the firing rate and Fano factor. In all figures, we set  $\tau_b = \tau_s = 1.315/c$ ,  $r_b = 37c$ , and  $r_s = 3c$  which gives a Fano factor of  $F_{\text{in}} = 20.0017 \approx 20$  for any value of  $c$  (from Eq. (11)). Changing  $c$  effectively scales the timescale of presynaptic spiking, hence scaling  $r_{\text{in}}$ , without changing  $F_{\text{in}}$ .

## 3 Results

We analyze the synaptic response to different patterns of presynaptic inputs using a stochastic model of short term synaptic depression in which a presynaptic neuron makes  $M$  functional contacts onto a postsynaptic neuron [60,22,24,19]. The input to the presynaptic neuron is a spike train denoted by  $I(t)$ . Neurotransmitter vesicles are released probabilistically in response to each presynaptic spike. Specifically, a contact with a readily available vesicle releases this vesicle with probability  $p_r$  in response to a single presynaptic spike. After a synaptic contact has released its neurotransmitter vesicle, it enters a refractory state where it is unable to release again until the vesicle is replaced. The duration of this



**Fig. 1 Rate of vesicle release as a function of presynaptic firing rate for various presynaptic Fano factors.** The rate of vesicle release,  $r_x$ , is an increasing function of presynaptic firing rate,  $r_{\text{in}}$ . Vesicle release is slower for the irregular spiking model than for the Poisson and regular spiking model.

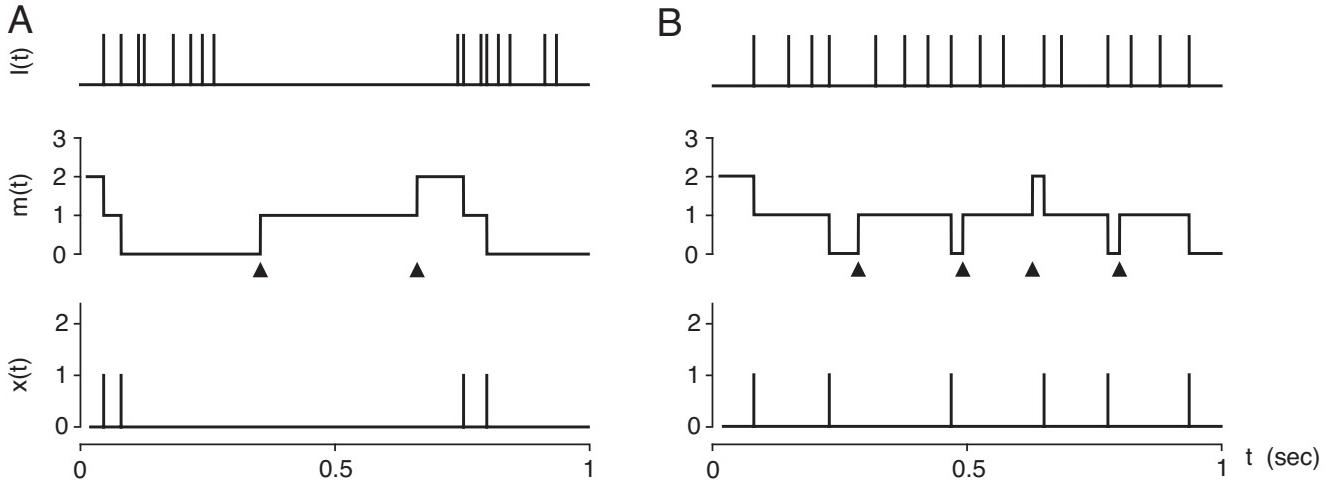
refractory period is an exponentially distributed random variable with mean  $\tau_u$ , so that vesicle recovery is Poisson in nature.

We are interested in how the statistics of the presynaptic spike train determine the statistics of the synaptic response. The presynaptic statistics are quantified using the presynaptic firing rate,  $r_{\text{in}}$ , the presynaptic auto-covariance function,  $R_{\text{in}}(\tau)$ , and the Fano factor,  $F_{\text{in}}(T)$ , of the number of presynaptic spikes during a window of length  $T$ . Similarly, we quantify the statistics of the synaptic response using the mean rate of vesicle release,  $r_x$ , the auto-covariance of vesicle release,  $R_x(\tau)$ , and the Fano factor,  $F_x(T)$ , of the number of vesicles released during a window of length  $T$ . We will especially focus on Fano factors over large time windows and define  $F_{\text{in}} = \lim_{T \rightarrow \infty} F_{\text{in}}(T)$ ,  $F_x = \lim_{T \rightarrow \infty} F_x(T)$  accordingly. See Methods for more details.

We begin by considering the effect of  $F_{\text{in}}$  on the mean rate of vesicle release,  $r_x$ . We then examine the dependence of  $F_x(T)$  on the length,  $T$ , of the time window over which vesicle release events are counted. Finally, we show that short term synaptic depression promotes Poisson-like responses to non-Poisson presynaptic inputs.

### 3.1 Irregular presynaptic spiking reduces the rate at which neurotransmitter vesicles are released

We first briefly investigate the dependence of the rate of vesicle release,  $r_x$ , on the rate and variability of the presynaptic spike train, as measured by  $r_{\text{in}}$  and  $F_{\text{in}}$  re-



**Fig. 2** Synaptic response to an irregular and a regular presynaptic spike train. **A)** An irregular spike train,  $I(t)$ , drives a depressing synapse. Each vesicle release event is indicated by a vertical bar with height indicating the number of vesicles released (here, all events release just one vesicle). Each time a vesicle is released, the number of available vesicles,  $m(t)$ , is decremented accordingly. Vesicle recovery increments  $m(t)$  and occurs randomly in time (vesicle recovery events indicated by filled triangles). **B)** Same as (A) except for a more regular presynaptic spike train. Note that, even though the same number of spikes occur in (A) and (B), the irregular spike train is less effective in releasing vesicles. This occurs because all vesicles are depleted by the first few spikes in a burst and subsequent spikes in that burst are unable to release vesicles. For illustrative purposes, we set  $M = 3$  in this figure.

spectively. Vesicle release rate generally increases with  $r_{in}$ , but saturates to  $r_x = M/\tau_u$  whenever  $p_r r_{in} \gg 1/\tau_u$  since synapses are depleted in this regime (Fig. 1).

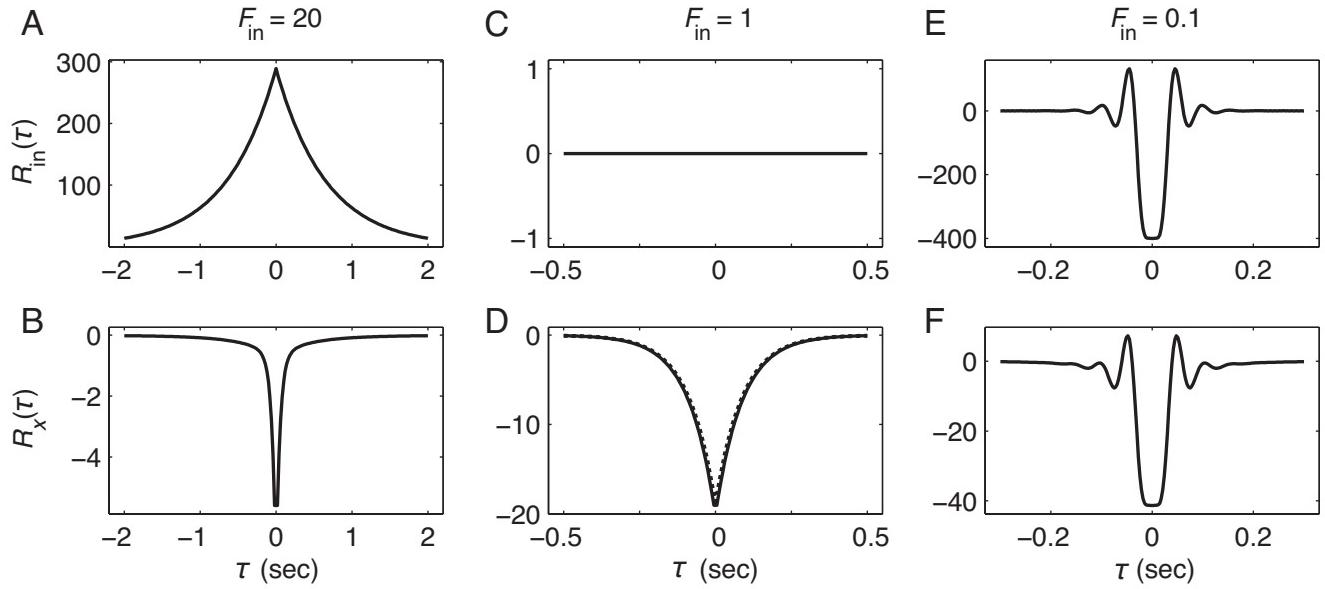
When presynaptic spike times occur as a Poisson process (so that  $F_{in} = 1$ ), the mean rate of vesicle release is given by  $r_x = Mp_r r_{in}/(p_r r_{in}\tau_u + 1)$  [22, 19, 43]. Interestingly, vesicle release is slower for more irregular presynaptic spiking and faster for more regular presynaptic spiking even when presynaptic spikes arrive at the same mean rate (Fig. 1, also see [18]). This can be understood by noting that, for the irregular input model, spikes arrive in bursts of higher firing rate followed by durations of lower firing rate. Vesicles are depleted by the first few spikes in a burst and subsequent spikes in that burst are ineffective and therefore essentially “wasted” spikes (Fig. 2A). When presynaptic spikes arrive more regularly, more vesicles are released on average (Fig. 2B).

### 3.2 Variability in the number of vesicles released in a time window decreases with window size

We now consider how the variability of the synaptic response to a presynaptic input depends on the timescale over which this variability is measured. We quantify the variability of the synaptic response using the Fano factor,  $F_x(T)$ , which is defined to be the variance-to-mean ratio of the number of vesicles released in a time window of length  $T$  (see Methods)

and can be calculated from an integral of the auto-covariance function,  $R_x(\tau)$ , using Eq. (3).

The auto-covariance of a Poisson presynaptic spike train is simply a delta function at the origin,  $R_{in}(\tau) = r_{in}\delta(\tau)$ , and the Fano factor over any window size is therefore equal to one,  $F_{in}(T) = 1$  (Figs. 3C and 4C). The auto-covariance of the synaptic response when presynaptic inputs are Poisson consists of a delta function at the origin surrounded by a double-sided exponential with a negative peak (see Eq. (4) and Fig. 3D) that decays with timescale  $\tau_0 = \tau_u/(1 + p_r r_{in}\tau_u)$ . The fact that the auto-covariance is negative away from  $\tau = 0$  implies that the Fano factor,  $F_x(T)$ , is monotonically decreasing in the window size,  $T$  (see Eq. (7) and Fig. 4D). For small  $T$ , the mass of the delta function at the origin dominates the integral in Eq. (3) so that the Fano factor is approximately equal to the ratio of this mass to the mean rate,  $r_x$ , at which vesicles are released. As  $T$  increases, the negative mass of the exponential peak subtracts from the positive contribution of the delta function and decreases the Fano factor. In particular,  $F_x(T) \approx D - ET + \mathcal{O}(T^2)$  where  $Dr_x$  is the mass of the delta function in  $R_x(\tau)$  and  $-Er_x$  is the peak of the exponential in  $R_x(\tau)$  (see Eqs. (5) and (6)). As  $T$  continues to increase,  $F_x(T)$  monotonically decreases towards its limit,  $F_x := \lim_{T \rightarrow \infty} F_x(T) = D - 2E\tau_0$ . Thus, short term synaptic depression converts a Fano factor that is constant with respect to window size into one that decreases with window size (Fig. 4C,D).



**Fig. 3 Presynaptic auto-covariance functions and auto-covariance of vesicle release for three input models.** Auto-covariance functions of presynaptic spike trains (top row) and the synaptic response they evoke (bottom row) for three different presynaptic input models: **A,B**)irregular ( $F_{\text{in}} = 20$ ), **C,D**)Poisson ( $F_{\text{in}} = 1$ ) and **E,F**) regular ( $F_{\text{in}} = 0.1$ ). Each auto-covariance function has a Dirac delta function at the origin that is not depicted here. Dotted line in (D) is from the approximation in Eq. (4) and solid line is from exact calculation obtained using numerics for the regular input model with  $\theta = 1$  (see Methods), but the two are nearly indistinguishable.  $R_{\text{in}}(\tau)$  has units (spikes/sec)<sup>2</sup> and  $R_x(\tau)$  has units (vesicles/sec)<sup>2</sup>. Short term depression introduces negative temporal correlations even when presynaptic spike trains are temporally uncorrelated (C,D) or positively correlated (A,B).

When presynaptic spike times are not Poisson, the statistics of the postsynaptic response cannot be derived analytically using the methods utilized for the Poisson input model. Instead, we use the fact that the synapse model can be represented using a continuous time Markov chain, which can be analyzed to derive expressions for the response statistics in terms of an infinitesimal generator matrix (see Methods).

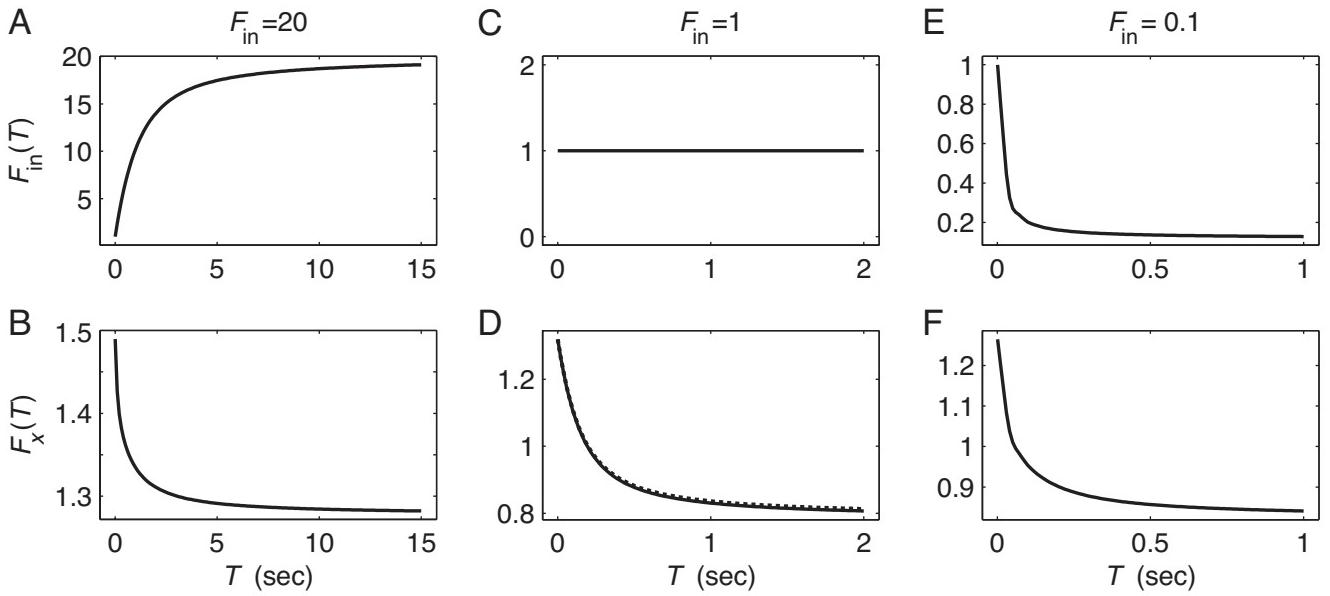
Irregular presynaptic spiking (i.e., inputs with  $F_{\text{in}} > 1$ ) is achieved by varying the rate of presynaptic spiking randomly in time to produce a doubly stochastic Poisson process (see Methods). For this model, the input auto-covariance is a delta function at the origin surrounded by an exponential peak (see Eq. 10 and Fig. 3A). The input Fano factor therefore increases with window size (see Eq. 11 and Fig. 4A). The positive temporal correlations exhibited in the input auto-covariance function are canceled by the temporal de-correlating effects of short term synaptic depression [26, 25, 24]. For the parameters chosen in this study, this de-correlation outweighs the positive presynaptic correlations so that the auto-covariance function of the response is negative away from  $\tau = 0$  (Fig. 3B), although parameters can also be chosen so that temporal correlations in the response are small and positive [25]. As with the Poisson input model, negative temporal correlations cause the response Fano factor to decrease

with window size (Fig. 4B). Thus short term synaptic depression and stochastic vesicle dynamics can convert a presynaptic Fano factor that *increases* with window size into one that *decreases*.

Regular presynaptic spiking is achieved by generating a renewal process with gamma-distributed interspike-intervals. The input auto-covariance function for this model exhibits temporal oscillations (Eq. (15) and Fig. 3E) and the Fano factor generally decreases with window size (Fig. 4E). Perhaps unsurprisingly, the auto-covariance function of the synaptic response exhibits oscillations and the response Fano factor decreases with window size (Figs. 3F and 4F).

For all three input models, the variability of the synaptic response is larger over shorter time windows and smaller over larger time windows. A postsynaptic neuron that is in an excitable regime will generally respond most effectively to inputs that exhibit more variability over short time windows [46, 47, 39, 38]. In addition, rate coding is often more efficient when spike counts over larger time windows are less variable [64]. Thus, the dependence of  $F_x(T)$  on window size is especially efficient for the neural transmission of rate-coded information [24].

In addition to the temporal dependence of  $F_x(T)$  introduced by short term depression, note that the response Fano factor for the irregular input model is sub-



**Fig. 4 Presynaptic and response Fano factor as a function of window size for three input models.** Presynaptic and response Fano factors,  $F_{\text{in}}(T)$  and  $F_x(T)$ , as a function of the window size over which inputs or vesicles are counted (see Methods), obtained by applying Eq. (1) to the auto-covariance functions in Fig. 3. Short term depression causes response Fano factor to decrease with window size even when presynaptic Fano factor increases with window size (A,B) or is independent of window size (C,D). Also, response Fano factors are near 1 even when presynaptic Fano factors are not (B and F).

stantially smaller than the input Fano factor (Fig. 4). Conversely, the response Fano factor for the regular input model is larger than the input Fano factor (Fig. 4). For both models, the response Fano factor is substantially nearer to 1 than the input Fano factor. We explain this phenomenon next.

### 3.3 Depleted synapses exhibit Poisson-like variability even when presynaptic inputs are highly non-Poisson

We now investigate the dependence of the variability in synaptic response on the rate and variability of the presynaptic input. Since we have already discussed the dependence of  $F_x(T)$  on  $T$  above, we will focus here on the Fano factor calculated over long time windows,  $F_x = \lim_{T \rightarrow \infty} F_x(T)$ .

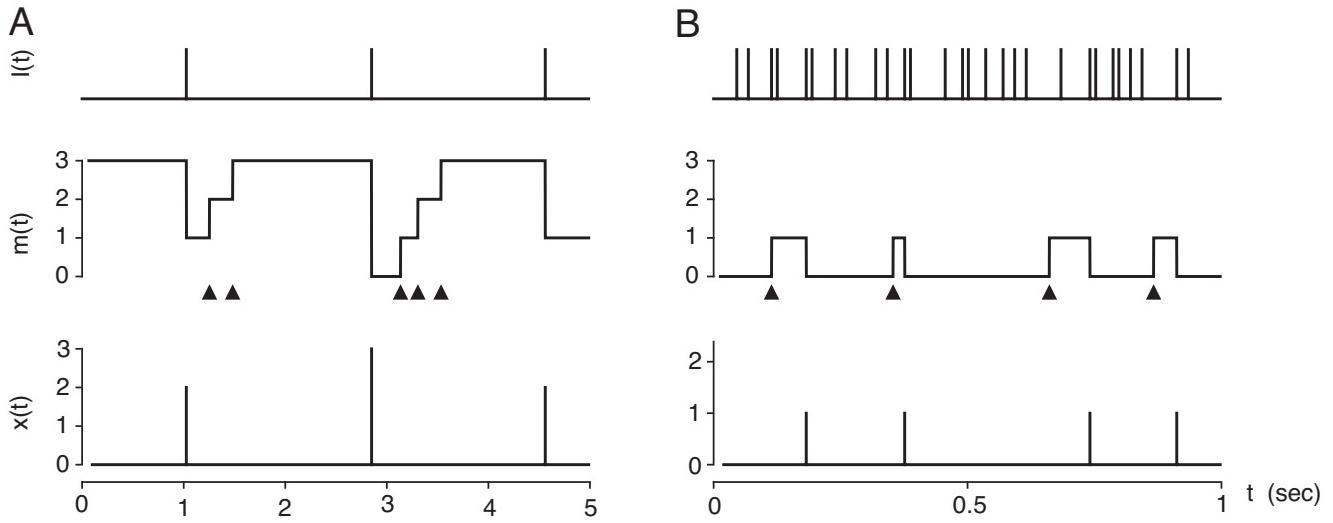
We first consider parameter regimes where the effective rate of presynaptic inputs is much slower than the rate of vesicle recovery ( $p_r r_{\text{in}} \ll 1/\tau_u$ ). In such a regime, each contact is likely to recover between two consecutive presynaptic spikes and therefore all  $M$  contacts are likely to have a vesicle ready to release when each spike arrives (Fig. 6A). In this limit, the number of vesicles released by each spike is an independent binomial variable with mean  $\langle w_j \rangle = p_r M$  and variance  $\text{var}(w_j) = M p_r (1 - p_r)$ . The number,  $N_x(T)$ , of vesicles released in a time window of length  $T$  can then be represented as a sum of  $N_{\text{in}}(T)$  independent binomial random variables (i.e., a random sum). The mean of

this sum is given by  $\langle N_x(T) \rangle = \langle N_{\text{in}}(T) \rangle \langle w_j \rangle$ , which implies that  $r_x = M p_r r_{\text{in}}$  in this limit. Similarly, the variance of this sum is given by [31]  $\text{var}(N_x(T)) = \langle N_{\text{in}}(T) \rangle \text{var}(w_j) + \langle w_j \rangle^2 \text{var}(N_{\text{in}}(T))$ , which implies

$$\begin{aligned} \lim_{r_{\text{in}} \rightarrow 0} F_x(T) &= \langle w_j \rangle F_{\text{in}}(T) + \text{var}(w_j)/\langle w_j \rangle \\ &= 1 + p_r(M F_{\text{in}}(T) - 1). \end{aligned} \quad (16)$$

Eq. (16) is verified for the Poisson input model by taking  $r_{\text{in}} \rightarrow 0$  in Eqs. (7). For the irregular and regular input models, Eq. (16) should be interpreted heuristically, as it was derived heuristically. A counterexample to Eq. (16) for the irregular input model can be constructed by fixing  $r_f$  and  $\tau_f$ , then letting  $\tau_s \rightarrow \infty$  and  $r_s \rightarrow 0$  to achieve the  $r_{\text{in}} \rightarrow 0$  limit. In this case, our assumption that each contact is increasingly likely to recover between two consecutive spikes is violated and Eq. (16) is not valid (not pictured). Regardless, we verify numerically that Eq. (16) is accurate when  $r_{\text{in}}$  is decreased toward zero while keeping  $F_{\text{in}}$  fixed (Fig. 5).

We now discuss the statistics of the postsynaptic response when the effective presynaptic spiking is much faster than vesicle recovery ( $p_r r_{\text{in}} \gg 1/\tau_u$ ). In such a regime, incoming spikes occur much more frequently than recovery events and synapses becomes depleted. As a result, the number of vesicles released over a long time window is determined predominantly by the number of recovery events in that time window and largely independent from the number of presynaptic spikes (Fig. 6B)



**Fig. 6 Vesicle release dynamics at low and high presynaptic firing rates.** **A)** At low presynaptic rates, vesicles are recovered (i.e.,  $m(t)$  returns to  $M = 3$ ) between presynaptic spikes. Thus, the number of vesicles released by each presynaptic spike is approximately an independent binomial random variable with mean  $p_{\text{pr}}M$  and variance  $p_{\text{pr}}M(1 - M)$ . **B)** At high presynaptic rates, vesicles are released almost immediately after they are recovered. Thus, the number of vesicles released over a time window of length  $T$  is approximately a Poisson random variable with mean and variance  $TM/\tau_u$ .

[19,43]. The synaptic response therefore inherits the Poisson statistics of the recovery events so that

$$\lim_{r_{\text{in}} \rightarrow \infty} F_x(T) = 1.$$

For the Poisson input model, this limit can be made more precise in the  $T \rightarrow \infty$  limit by expanding Eq. (8) in terms of the parameter  $\alpha = 1/(p_{\text{pr}}r_{\text{in}}\tau_u)$  to obtain

$$F_x = 1 - 2\alpha + 4\alpha^2 + \mathcal{O}(\alpha^3) \quad (17)$$

which converges to 1 as  $r_{\text{in}}\tau_u \rightarrow \infty$ . For the irregular and regular input models, we verify in Fig. 5 that  $F_x \rightarrow 1$  when  $r_{\text{in}}$  is increased while keeping  $F_{\text{in}}$  fixed.

The time constant,  $\tau_u$ , at which a synapse recovers from short term depression has been measured in a number of experimental studies and is often found to be several hundred milliseconds [57,59,34,23,22,28,41]. Therefore, for even moderate presynaptic firing rates, synapses are often in a highly depleted state. As discussed above, this promotes Poisson-like variability in the synaptic response. This provides one possible mechanism through which irregular Poisson-like firing can be sustained in neuronal populations [11].

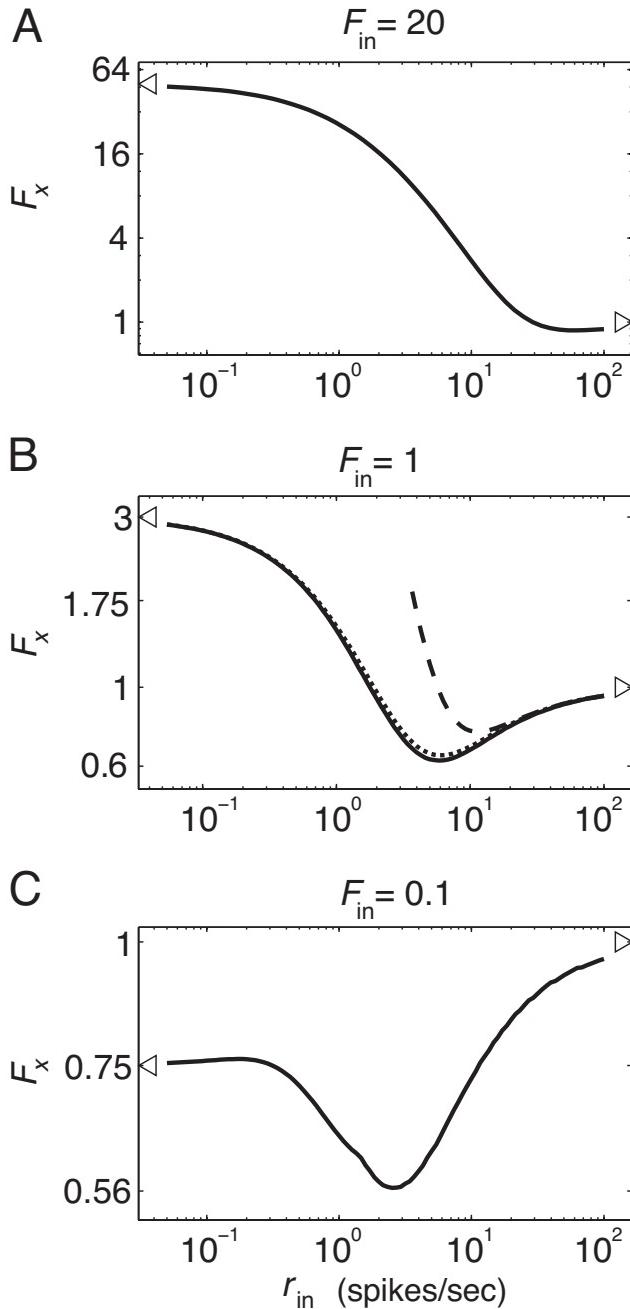
## 4 Discussion

We used continuous time Markov chain methods to derive the response statistics of a stochastic model of short term synaptic depression with three different presynaptic input models. We then used this analysis to understand how the mean presynaptic firing rate and the

variability of presynaptic spiking interact with synaptic dynamics to determine the mean rate of vesicle release and variability in the number of vesicles released. This analysis revealed a number of fundamental, qualitative dependencies of response statistics on presynaptic spiking statistics. Some of the dependencies have been previously noted in the literature and some have not.

The number of vesicles released over a time window is smaller for irregular inputs than for more regular inputs (Figs. 1 and 2) given the same number of presynaptic spikes. Thus, regular presynaptic spiking is more efficient at driving synapses. This mechanism competes with a well-known property of excitable cells: that they are driven more effectively by irregular, positively correlated synaptic input currents [46,47,39,38]. In addition, a *population* of presynaptic spike trains drives a postsynaptic neuron more efficiently when the population-level activity is more irregular, for example due to pairwise correlations [19]. Together, these results suggest that a postsynaptic neuron is most efficiently driven by presynaptic populations that exhibit small or negative auto-correlations, but positive pairwise cross-correlations.

Our model predicts that the de-correlating effects of short term depression and stochastic vesicle dynamics can produce negative temporal auto-correlations in the synaptic response even when presynaptic spiking is temporally uncorrelated or positively correlated, in agreement with previous studies [25,19]. This yields a response Fano factor that decreases with window size, as observed in some recorded data [30]. We note, though,



**Fig. 5 Response Fano factor as a function of presynaptic firing rate for three input models.** Response Fano factors calculated over large windows for **A)** the irregular input model **B)** Poisson input model and **C)** regular input model. Fano factors approach 1 at high presynaptic firing rates regardless of the presynaptic Fano factor (triangle on right is placed at  $F_x = 1$ ). At low presynaptic firing rates, response Fano factors approach the value given in Eq. (16) (indicated by triangle on left). Dotted line in (B) is from closed form approximation in Eq. (8) and dashed line is from the expansion given in Eq. (17).

that some parameter choices can yield positively correlated synaptic response when presynaptic inputs are positively correlated [25] and neuronal membrane dynamics can introduce positive correlations to a postsynaptic spiking response even when synaptic currents are not positively correlated in time [37]. This is consistent with several studies showing positive temporal correlations in recorded spike trains [4, 15, 3, 11].

We predict that moderate or high firing rates can induce a Poisson-like synaptic response even when presynaptic inputs are non-Poisson (Fig. 5 and [18]). This is because even moderate firing rates can deplete synapses and depleted synapses inherit the Poisson-like variability of synaptic vesicle recovery (Fig. 6B and [19, 43]). At lower firing rates, short term depression and synaptic variability can increase or decrease Fano factor. For example, in Fig. 5B, the response Fano factor is larger than the presynaptic Fano factor ( $F_{in} = 1$ ) at low firing rates, decreases at higher firing rates, then approaches  $F_x = 1$  at higher firing rates. This complex dependence of firing rate on Fano factor might be related to the stimulus dependence of Fano factors observed in several cortical brain regions [11].

Our conclusion that fast presynaptic spiking causes Poisson-like variability in the synaptic response relied on the assumption that vesicle recovery times are exponentially distributed. The exponential distribution is a justifiable choice for recovery times only if recovery times obey a memoryless property: having already waited  $t$  units of time for a recovery event, the probability of waiting an additional  $s$  units of time does not depend on  $t$ . The precise mechanics of vesicle re-uptake and docking determine whether this is an appropriate assumption. If recovery times have a different probability distribution, then the synaptic response will inherit the properties of this distribution at high presynaptic firing rates instead of inheriting the Poisson-like nature of exponentially distributed recovery times.

Previous methods have been developed to analyze the synaptic response of the model used here. In [18, 24], the model restricted to the  $M = 1$  case is analyzed for presynaptic spike trains that are renewal processes. This includes the Poisson and the regular input model discussed here, but excludes the irregular input model in which the spike train is a non-renewal inhomogeneous Poisson process. In [43] approximations are obtained for the case where the presynaptic spike train is an inhomogeneous Poisson process, but the approximation is only valid when the rate-modulation of the Poisson process is small compared to the average firing rate. Thus, these approximations are only valid for the irregular input model when  $r_f - r_s \ll r_s$ . Other studies [32, 36] use a deterministic synapse model that

implicitly treats the number of available vesicles as a continuous rather than a discrete quantity. This deterministic model represents the trial average of the model considered here and can vastly underestimate the variability of a synaptic response [43].

A more detailed synapse model allows for multiple docking sites at a single contact [62, 19]. This model can yield different response properties than the model used here in certain parameter regimes [19]. Even though this more detailed model can be represented as a continuous-time Markov chain, the analysis of this model would be significantly more complex than the analysis considered here since it would be necessary to keep track of the number of readily releasable vesicles at each contact separately. This would result in a Markov chain with  $K \times N^M$  states where  $M$  is the number of contacts,  $N$  is the number of docking sites per contact and  $K$  is the number of states used for the presynaptic input model ( $K = 1$  for the Poisson input model,  $K = 2$  for the irregular input model, and  $K = \theta$  for the regular input model).

To quantify the synaptic response to a presynaptic spike train, we focused on the statistics of the number of vesicles released in a time window. Postsynaptic neurons observe changes in synaptic conductance in response to presynaptic spikes. The synaptic conductance are often modeled in such a way that they can be easily derived from our process  $x(t)$  through a convolution:  $g(t) = \int_0^t x(t-s)\alpha(s)ds$  where  $g(t)$  is the synaptic conductance elicited by a presynaptic spike train and  $\alpha(s)$  is a kernel representing the characteristic postsynaptic conductance elicited by the release of a single neurotransmitter vesicle. Since this mapping is linear, the statistics of  $g(t)$  can easily be derived in terms of the statistics of  $x(t)$  [53, 43].

## References

1. L F Abbott and Wade G Regehr. Synaptic computation. *Nature*, 431(7010):796–803, October 2004.
2. L F Abbott, J A Varela, K Sen, and S B Nelson. Synaptic depression and cortical gain control. *Science*, 275(5297):220–224, January 1997.
3. R. Baddeley, L.F. Abbott, M.C.A. Booth, F. Sengpiel, T. Freeman, E.A. Wakeman, and E.T. Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 264(1389):1775–1783, 1997.
4. W. Bair, C. Koch, W. Newsome, and K. Britten. Power spectrum analysis of bursting cells in area mt in the behaving monkey. *The Journal of neuroscience*, 14(5):2870–2892, 1994.
5. F. Barbieri and N. Brunel. Can attractor network models account for the statistics of firing during persistent activity in prefrontal cortex? *Front in Neurosci*, 2(1):114, 2008.
6. Tiago Branco and Kevin Staras. The probability of neurotransmitter release: variability and feedback control at single synapses. *Nature Rev Neurosci*, 10(5):373–383, May 2009.
7. P. C. Bressloff. Mean-field theory of globally coupled integrate-and-fire neural oscillators with dynamic synapses. *Phys. Rev. E*, 60:2160–2170, Aug 1999.
8. K.H. Britten, M.N. Shadlen, W.T. Newsome, J.A. Movshon, et al. Responses of neurons in macaque mt to stochastic motion signals. *Visual neuroscience*, 10:1157–1157, 1993.
9. G.T. Buracas, A.M. Zador, M.R. DeWeese, and T.D. Albright. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron*, 20(5):959–969, 1998.
10. F.S. Chance, S.B. Nelson, and L.F. Abbott. Synaptic depression and the temporal response characteristics of v1 cells. *J Neurosci*, 18(12):4785, 1998.
11. M.M. Churchland, M.Y. Byron, J.P. Cunningham, L.P. Sugrue, M.R. Cohen, G.S. Corrado, W.T. Newsome, A.M. Clark, P. Hosseini, B.B. Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369–378, 2010.
12. Daniel L Cook, Peter C Schwindt, Lucinda A Grande, and William J Spain. Synaptic depression in the localization of sound. *Nature*, 421(6918):66–70, January 2003.
13. D Cox. *Renewal Theory*. London: Methuen and Co, 1962.
14. D Cox and V Isham. *Point processes*. London: Chapman and Hall, 1980.
15. Y. Dan, J.J. Atick, and R.C. Reid. Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *The Journal of Neuroscience*, 16(10):3351–3362, 1996.
16. P. Dayan and L.F. Abbott. *Theoretical Neurosci: Computational and mathematical modeling of neural systems*. Taylor & Francis, 2001.
17. J de la Rocha and R Moreno. Correlations modulate the non-monotonic response of a neuron with short-term plasticity. *Neurocomputing*, 2004.
18. J de la Rocha and A Nevado. Information transmission by stochastic synapses with short-term depression: neural coding and optimization. *Neurocomputing*, 2002.
19. Jaime de la Rocha and Néstor Parga. Short-term synaptic depression causes a non-monotonic response to correlated stimuli. *J Neurosci*, 25(37):8416–8431, September 2005.
20. Jaime de la Rocha and Néstor Parga. Thalamocortical transformations of periodic stimuli: the effect of stimulus velocity and synaptic short-term depression in the vibrissa-barrel system. *J Comput Neurosci*, 25(1):122–140, August 2008.
21. A.A. Faisal, L.P.J. Selen, and D.M. Wolpert. Noise in the nervous system. *Nature Rev Neurosci*, 9(4):292–303, 2008.
22. G. Fuhrmann, I. Segev, H. Markram, and M. Tsodyks. Coding of temporal information by activity-dependent synapses. *J Neurophys*, 87(1):140, 2002.
23. M. Galarreta and S. Hestrin. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neurosci*, 1:587–594, 1998.
24. M.S. Goldman. Enhancement of information transmission efficiency by synaptic failures. *Neural Comput*, 16(6):1137–1162, 2004.
25. M.S. Goldman, P. Maldonado, and LF Abbott. Redundancy reduction and sustained firing with stochastic depressing synapses. *The Journal of Neuroscience*, 22(2):584–591, 2002.

26. M.S. Goldman, S.B. Nelson, and LF Abbott. Decorrelation of spike trains by synaptic depression. *Neurocomputing*, 26:147–153, 1999.
27. Lucinda A Grande and William J Spain. Synaptic depression as a timing device. *J Physiol*, 20:201–210, June 2005.
28. Jesse E Hanson and Dieter Jaeger. Short-term plasticity shapes the response to simulated normal and parkinsonian input patterns in the globus pallidus. *J Neurosci*, 22(12):5164–5172, June 2002.
29. A. Harsch and H.P.C. Robinson. Postsynaptic variability of firing in rat cortical neurons: the roles of input synchronization and synaptic nmda receptor conductance. *The Journal of Neuroscience*, 20(16):6181–6192, 2000.
30. P. Kara, P. Reinagel, and R.C. Reid. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron*, 27(3):635–646, 2000.
31. S. Karlin and H.M. Taylor. *A first course in stochastic processes*. Academic press, 1975.
32. B Lindner, D Gangloff, A Longtin, and J E Lewis. Broadband Coding with Dynamic Synapses. *J Neurosci*, 29(7):2076–2087, February 2009.
33. A Litwin-Kumar and B Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat Neurosci*, 2012.
34. H. Markram, Y. Wang, and M. Tsodyks. Differential signaling via the same axon of neocortical pyramidal neurons. *Proc Natl Acad Sci USA*, 95(9):5323, 1998.
35. C.J. McAdams and J.H.R. Maunsell. Effects of attention on the reliability of individual neurons in monkey visual cortex. *Neuron*, 23(4):765–773, 1999.
36. Matthias Merkel and Benjamin Lindner. Synaptic filtering of rate-coded information. *Phys Rev E*, 81(4), April 2010.
37. R. Moreno-Bote and N. Parga. Auto-and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Physical review letters*, 96(2):28101, 2006.
38. R. Moreno-Bote and N. Parga. Response of integrate-and-fire neurons to noisy inputs filtered by synapses with arbitrary timescales: Firing rate and correlations. *Neural Computation*, 22(6):1528–1572, 2010.
39. R. Moreno-Bote, A. Renart, and N. Parga. Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. *Neural Computation*, 20(7):1651–1705, 2008.
40. AM Oswald and NN Urban. Interactions between behaviorally relevant rhythms and synaptic plasticity alter coding in the piriform cortex. *J Neurosci (In Press)*, 2012.
41. M Rav-Acha, N Sagiv, I Segev, H Bergman, and Y Yarom. Dynamic and spatial features of the inhibitory pallidal GABAergic synapses. *J Neurosci*, 135(3):791–802, 2005.
42. R Rosenbaum. *The transfer and propagation of correlated neuronal activity*. PhD thesis, University of Houston, 2011.
43. R Rosenbaum, J Rubin, and B Doiron. Short term synaptic depression imposes a frequency dependent filter on synaptic information transfer. *Public Library of Sci Comput Biol*, 8(6), 2012.
44. Jason S Rothman, Laurence Cathala, Volker Steuber, and R Angus Silver. Synaptic depression enables neuronal gain control. *Nature*, 457(7232):1015–1018, February 2009.
45. Z Rotman, P Y Deng, and V A Klyachko. Short-Term Plasticity Optimizes Synaptic Information Transmission. *J Neurosci*, 31(41):14800–14809, October 2011.
46. E. Salinas and T.J. Sejnowski. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *Journal of Neuroscience*, 20(16):6193, 2000.
47. E. Salinas and T.J. Sejnowski. Integrate-and-fire neurons driven by correlated stochastic input. *Neural computation*, 14(9):2111–2155, 2002.
48. W Senn, H Markram, and M Tsodyks. An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Comput*, 13(1):35–67, January 2001.
49. M.N. Shadlen and W.T. Newsome. Noise, neural codes and cortical organization. *Findings and Current Opinion in Cognitive Neuroscience*, 4:569–579, 1998.
50. M.N. Shadlen and W.T. Newsome. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896, 1998.
51. W.R. Softky and C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *Journal of Neuroscience*, 13(1):334–350, 1993.
52. C.F. Stevens, A.M. Zador, et al. Input synchrony and the irregular firing of cortical neurons. *Nature neuroscience*, 1(3):210–217, 1998.
53. T. Tetzlaff, S. Rotter, E. Stark, M. Abeles, A. Aertsen, and M. Diesmann. Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. *Neural Comput*, 20(9):2133–2184, 2008.
54. DJ Tolhurst, JA Movshon, and AF Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision research*, 23(8):775–785, 1983.
55. M Tsodyks, K Pawelzik, and H Markram. Neural networks with dynamic synapses. *Neural Comput*, 10(4):821–835, May 1998.
56. M. Tsodyks, A. Uziel, H. Markram, et al. Synchrony generation in recurrent networks with frequency-dependent synapses. *J Neurosci*, 20(1):825–835, 2000.
57. M V Tsodyks and H Markram. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci USA*, 94(2):719–723, January 1997.
58. C. Van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.
59. J A Varela, K Sen, J Gibson, J Fost, L F Abbott, and S B Nelson. A quantitative description of short-term plasticity at excitatory synapses in layer 2/3 of rat primary visual cortex. *J Neurosci*, 17(20):7926–7940, October 1997.
60. D. Vere-Jones. Simple stochastic models for the release of quanta of transmitter from a nerve terminal. *Australian & New Zealand Journal of Statistics*, 8(2):53–63, 1966.
61. C. Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10(6):1321–1371, 1998.
62. X J Wang. Fast burst firing and short-term synaptic plasticity: a model of neocortical chattering neurons. *J Neurosci*, 89(2):347–362, 1999.
63. A. Zador. Impact of synaptic unreliability on the information transmitted by spiking neurons. *J Neurophys*, 79(3):1219, 1998.
64. E Zohary, MN Shadlen, and WT Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140–3, Jul 1994.
65. R.S. Zucker and W.G. Regehr. Short-term synaptic plasticity. *Annual Rev of Phys*, 64(1):355–405, 2002.